DOCUMENT RESUME

ED 241 577 TM 840 117

TITLE	Test Validation for 12,000 Jobs: An Application of Job Classification and Validity Generalization Analysis to the General Aptitude Test Battery.
INSTITUTION	Employment and Training Administration (DOL), Washington, DC. Div. of Counseling and Test Development.
REPORT NO	USES-TRR-45
PUB DATE	83
NOTE	60p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Adults; Career Counseling; Cognitive Ability; *Federal Programs; *Job Analysis; Longitudinal Studies; Perception; *Predictor Variables; Psychomotor Skills; Test Construction; *Testing Programs; *Test Validity
IDENTIFIERS	*General Aptitude Test Battery; Validity Research

ABSTRACT

This report cumulates the results of 515 validation studies carried out over a 45-year period by the United States Employment Service, and relates these findings to five systems of job classification and job analysis. Correction for sampling error shows that general cognitive, perceptual, and psychomotor ability are valid predictors of job proficiency for all jobs, though there is considerable variation in validity across jobs. Correction for sampling error shows that cognitive and perceptual ability are valid predictors of training success for all jobs and that psychomotor ability is a valid predictor for all but a few high-complexity jobs. The relevant information in each of the five job analysis systems turned out to be the same dimension: job complexity. This dimension has been assessed for all 12,000 jobs in the "Dictionary of Occupational Titles" and the validity generalization analysis performed here thus extends to all jobs in the current volume. Cognitive ability increases in validity as job complexity increases while psychomotor ability increases in validity as complexity decreases. Thus a shift in weight from cognitive ability to psychomotor ability across categories of job complexity produces average multivariate validity ranging from .49 to .59 for job proficiency and from .59 to .65 for training success. (Author)

USES TEST RESEARCH REPORT NO. 45

TEST VALIDATION FOR 12,000 JOBS: AN APPLICATION OF JOB CLASSIFICATION AND VALIDITY GENERALIZATION ANALYSIS TO THE GENERAL APTITUDE TEST BATTERY

DIVISION OF COUNSELING AND TEST DEVELOPMENT
EMPLOYMENT AND TRAINING ADMINISTRATION
U. S. DEPARTMENT OF LABOR
WASHINGTON, D.C. 20213

1983

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENT

The United States Employment Service (USES) conducts a test research program for developing testing tools useful in vocational counseling and placement.

The purpose of this series of reports is to provide results of significant test research projects as they are completed. These reports will be of interest to users of USES tests and to test research personnel in state agencies and other organizations.

This report cumulates the findings of validity studies done by hundreds of analysts working for the U.S. Employment Service over a 45-year span. Special thanks go to Ron Boese of the North Carolina Employment Security Commission for the computer analysis he conducted.

This report was written by Dr. John E. Hunter, Michigan State University, under contract to the Southern Test Development Field Center, North Carolina Employment Security Commission, Raleigh, North Carolina. The report was prepared for printing by staff of the Western Test Development Field Center, Utah Department of Employment Security.

Researchers are encouraged to express fully their professional judgment. Therefore, points of view or opinions stated in this document do not necessarily represent the official position or policy of the Department of Labor nor of the North Carolina Employment Security Commission.

ABSTRACT

This report cumulates the results of 515 validation studies carried out over a 45-year period by the U.S. Employment Service, and relates these findings to five systems of job classification and job analysis. Correction for sampling error shows that general cognitive, perceptual, and psychomotor ability are valid predictors of job proficiency for all jobs, though there is considerable variation in validity across jobs. Correction for sampling error shows that cognitive and perceptual ability are valid predictors of training success for all jobs and that psychomotor ability is a valid predictor for all but a few high-complexity jobs. The relevant information in each of the five job analysis systems turned out to be the same dimension: job complexity. This dimension has been assessed for all 12,000 jobs in the Dictionary of Occupational Titles (DOT; U.S. Department of Labor, 1977) and the validity generalization analysis performed here thus extends to all jobs in the current volume. Cognitive ability increases in validity as job complexity increases while psychomotor ability increases in validity as complexity decreases. Thus a shift in weight from cognitive ability to psychomotor ability across categories of job complexity produces average multivariate validity ranging from .49 to .59 for job proficiency and from .59 to .65 for training success.

INTRODUCTION

This report presents a basis for the validation of ability tests for all of the 12,000 jobs included in the Dictionary of Occupational Titles (DOT; U.S. Department of Labor, 1977). The empirical basis for this report is the 40 years of test validation carried out on the General Aptitude Test Battery (GATB) by the U.S. Employment Service. Results given here cumulate across 515 validation studies. The conceptual basis for the report is 40 years of work on job analysis as it pertains to aptitude requirements carried out largely within the Department of Labor, although often with impetus from small-scale studies done in industrial settings outside the government.

The research proceeded in two stages. First, five systems of job classification were compared for their capacity to predict the correlation between cognitive, perceptual, and psychomotor abilities and job performance. This comparison showed the relevant dimension to be job complexity. Second, job complexity was used to define five job families which include all the jobs in the DOT. Validity generalization techniques were then applied to each of these job families.

Any job classification scheme attempts to partition jobs into categories which will be useful for various purposes. In order to be useful in personnel selection research, a classification scheme must be related to the aptitude requirements of a job. That is, a classification scheme should break jobs into job families in such a way that there are differences in aptitude validity between categories, (cf. Pearlman, Schmidt, & Hunter, in press). In order to "validate" a job classification scheme, the data must show that validity varies in the predicted way across categories. The 515 validation studies conducted by the U.S. Employment Service can be used to evaluate any job analysis method which has been applied to the jobs in the DOT. Five methods are evaluated in this report. All were shown to be valid. In fact, all were shown to use essentially the same dimension to predict aptitude validity--job complexity. Thus, there was no new information to be gained by combining across job analysis procedures. The final output of this comparison was a set of five job families which cover the entire job spectrum, each of which has over 20 validation studies, and which span the job analysis spectrum in terms of job complexity. These families are essentially a synthesis of the valid aspects of the whole set of job analyses.

The 515 jobs in the U.S. Employment Service data base can be regarded as a sample of jobs from the 12,000 jobs of the DOT. Since the job complexity dimension used here has been applied to the entire DOT, results from the sample of 515 validation studies can be generalized to the corresponding stratification of the entire job population. Validity generalization analysis showed that this job family system provides a more than adequate basis for test validation for the DOT job population. There is variation within families, but even the worst case analysis revealed an average validity of

.32 for job proficiency and .37 for training success. Thus, even in the worst case, there would be very large gains in workforce productivity from using tests for selection.

As job complexity decreases, the validity of cognitive ability decreases while the validity of psychomotor ability increases. If an optimal combination of these two abilities is used for each job category, then there is a substantial improvement in overall prediction. The average composite validity ranges from .49 in the lowest complexity category to .59 in the highest complexity category. Thus, optimal use of the job complexity analysis available in the DOT provides an average baseline validity of .53 for the average job in the economy. Local validation studies need only be done if it is believed that it is possible to improve on that mark.

The Myth of the Invalid Test

Hunter (1980), Pearlman (1982), and Schmidt, Hunter and Pearlman (1981) have concluded from their work that validity generalization is very robust. In the Pearlman (1982), and Schmidt, Hunter, and Pearlman (1981) studies, which covered wide ranges of jobs, there were no jobs for which the major cognitive and perceptual tests were found invalid. (This conclusion was upheld for job knowledge and information tests as well.) The notion of the invalid test appears to be a myth based on illusions created by sampling error. This report will extend this conclusion in two ways: (1) by broadening the job base to the entire job spectrum, and (2) by extending the conclusion beyond cognitive abilities to psychomotor ability.

While there is no job where cognitive ability is invalid, there are jobs for which the validity of cognitive ability is low. Hunter (1980) noted that in the Helme, Gibson, and Brogden (Note 2) Army data, the validity tended to be high if the job involved decision-making or trouble-shooting, but tended to be low if the job was largely confined to carrying out a preset sequence of instruction.

It was hoped that in the present study the job analysis methods would identify the sources of variation in the validity of cognitive ability. If the job analysis could locate the jobs where the validity of cognitive ability is low, then other predictors could be sought for such jobs. In fact, the findings will show that the dimension of job complexity satisfies that hope. Cognitive ability has its lowest validity on jobs of low complexity. On those same jobs, psychomotor ability has its highest validity. Thus, as job complexity varies, it is possible to obtain high validity by shifting from cognitive to psychomotor ability as the prime predictor.

Multivariate Prediction: Promises and Limits

If the validity of any given ability varies across jobs, then we might hope that different abilities are maximally valid for different jobs. That is, if different abilities are highly valid in predicting performance in different jobs, then we could compensate for poor prediction with one ability by using some other ability which has high validity for that job. The General Aptitude Test Battery was constructed with such compensation in mind. The GATB measures nine aptitudes which were thought to be differentially valid across jobs. For example, manual dexterity might be valid for a job where quantitative aptitude is not valid. If different abilities are valid for different jobs, then it is possible to obtain high validity for all jobs by using a different multiple regression equation for each job. This regression equation would give high weight to that ability which has high validity for that job.

There are two limitations on the multiple regression strategy. First, some aptitudes vary together in their validity across jobs. If two aptitudes are always high or low together, then there is no gain from substituting one for the other in prediction. Second, unless the data for a single job are based on a very large sample, it is not possible to determine the correct regression equation on that job with any accuracy. This section will review the solution to the problem of covariation in aptitude validity across jobs, and the next section will note a solution for the problem of sampling error.

Hunter (Note 4) has shown that if aptitudes covary in their validity across jobs, then it is because the specific factors for those aptitudes are not relevant to the validity of the aptitude. Rather, the validity is determined by the general ability which underlies the two aptitudes. For example, there is a high correlation between various psychomotor aptitudes, which shows that there is a general psychomotor ability which partly determines all specific psychomotor aptitudes. Differences between psychomotor aptitudes are determined by specific factors which vary from one aptitude to the next. If the specific factors were relevant to some jobs and not to others, then the validity for two psychomotor aptitudes would vary differentially, as one specific factor is relevant to one job and the other specific factor is relevant to another job. However, if finger dexterity and manual dexterity always have either high validity or low validity together, then that fact shows that it is not the specific factors which are relevant to job performance but the general psychomotor factor which underlies the specific aptitudes.

If specific aptitudes covary perfectly in their validity across jobs, then it is the general factor underlying the specific aptitudes that is predicting job performance. Under these conditions, the test battery should be rescored in terms of general abilities rather than specific aptitudes. The general ability will have higher validity than either specific aptitude separately and will be correspondingly less subject to sampling error in the estimate of its validity from validation studies.

The first evidence for such covariation in aptitude validity across jobs was presented by Schmidt and Hunter (1978). They found that cognitive aptitudes varied together across jobs. For example, once sampling error was eliminated, verbal aptitude and quantitative aptitude had perfectly covarying validity across jobs; either both were high or both were low on each job. The same perfect covariation was found among perceptual aptitudes. Even the cognitive and perceptual aptitudes did not vary independently. The cross-correlation between cognitive and perceptual aptitude validity averaged .65.

Hunter (Note 4) correlated validity coefficients across jobs using the U.S. Employment Service data base of 515 jobs. He too found essentially perfect correlations among the cognitive aptitudes, among the perceptual aptitudes, and among the psychomotor aptitudes. Thus, Hunter concluded that cognitive, perceptual, and psychomotor ability should be scored as three-aptitude composite measures; there is little gain in considering the nine GATB aptitudes individually. Furthermore, Hunter found that if the abilities were perfectly measured, perceptual ability could be almost perfectly predicted from cognitive and psychomotor ability. Thus, if the abilities were perfectly measured, then perceptual ability would be a redundant predictor. Even for imperfectly measured abilities, perceptual ability has a multiple correlation of .80 when regressed onto the other abilities and hence is likely to contribute only rarely to an increase in predictive power.

However, while Hunter (Note 4) found high correlation between the cognitive- and perceptual aptitude validities, he found only low correlation between cognitive and psychomotor validities. Thus, there is reason to believe that cognitive and psychomotor abilities are complementary: if one has low validity, the other will often have high validity. Therefore, these two abilities often can profitably be substituted for each other in prediction. That is, validity for many jobs can be maximized by assigning larger weight to the more valid of these two abilities. This report will focus on aptitude composites, and will show that there is often substitution between cognitive and psychomotor ability in prediction.

Sampling Error and Job Families

The bane of multivariate prediction is sampling error. In order to have an adequate base for prediction from nine such highly correlated aptitudes as those on the GATB, one would need at least 1,000 persons in the validation study. But, as is typical of local validation studies, the U.S. Employment Service has rarely been able to obtain even as many as 200 persons for a given study. The average sample size for the 515 U.S. Employment Service studies is 75; Lent, Aurbach, and Levin (1971) similarly found the average for the field as a whole to be only 68.

The sample size requirements are diminished somewhat if aptitude composites are used for prediction. There are only three composites and they are less highly correlated. However, even a sample of 500 persons is rarely available.

There is but one solution to the problem of sampling error: One must form job families and validate using the cumulative methods of validity generalization (Callender, & Osburn, 1980; Hunter, Schmidt, & Jackson, 1982; Pearlman, Schmidt, & Hunter, 1980; Schmidt, Gast-Rosenberg, & Hunter, 1980; Schmidt, & Hunter, 1977; Schmidt, Hunter, & Caplan, 1981; Schmidt, Hunter, & Pearlman, in press; Schmidt, Hunter, Pearlman, & Shane, 1979; Hunter, Note 3). This solution has an additional advantage, i.e., the validity findings can be extended to other jobs in the same family without further research. Thus, a successful set of job families provides not only a large cumulative data base for multivariate prediction, but a basis for test validation for the entire job spectrum.

The present report analyzes job classification schemes which encompass the entire job spectrum, of which the 515 validation studies of the U.S. Employment Service are a large representative sample. Thus, the evaluation of any given job family approach extends to the entire set of jobs in the DOT.

Four bases for the formation of job families will be considered in detail: (1) test development analysts' judgments as to required aptitudes, (2) analysts' estimates of mean aptitudes levels, (3) the Data-People-Things hierarchy of worker functions (Fine, 1955), and (4) the U.S. Department of Labor (1979) Interest Guide Groups or Occupational Aptitude Patterns (OAPs). The Position Analysis Questionnaire (PAQ) of McCormick, Jeanneret, and Mecham (1972) will also be discussed.

Of the 515 validation studies, 425 used a criterion of job proficiency, while 90 used a criterion of training success. The job analysis findings presented in the next section are virtually identical for the whole data base and for proficiency or training success alone, and hence only the results for all studies are presented in the job analysis section that follows. Data are presented for proficiency and training criteria separately in the later section on validity generalization.

VALIDITY ACROSS JOB FAMILIES

This part of the report will present the analysis of the validity data for job classification systems. For each such system, there is a set of job families. These families are useful for purposes of personnel selection research to the extent that two things hold: (1) mean validity must vary across job families, and (2) the variance of validity within job families must be substantially lower than the variance of validity across the entire job spectrum (cf. Pearlman, Schmidt, & Hunter, in press). These properties are not independent but are mutually satisfied. The variance of validity within families is obtained by subtracting the variance of the means from the total variance. In the case of dimensional classification, these properties can be simultaneously measured by the correlation coefficient between dimension value and validity across jobs. That is, for each job there is a

dimension value and a validity coefficient. These can be correlated across jobs. If validity varies as a function of the dimension value, that is, if mean validity varies across job families, then the correlation will be high. If mean validity varies monotonically with the dimension value, then the ordinary Pearson product-moment correlation measures the usefulness of the classification dimension.

The correlations presented below are based on observed validity coefficients. They are thus guaranteed to be underestimates of the desired (true) correlations for job family dimensions because of the sampling error in the validity coefficients. That is, if a validity coefficient is computed on a small sample ($N < 2000$), then it will depart from the population value by a random amount that depends on the size of the sample. This random error produced by small sample size systematically reduces the size of the correlation between job dimension and validity. The extent of this reduction can be computed from known formulas for sampling error (Hunter, Schmidt, and Jackson, 1982; Hunter, Note 3). The correction is similar to the formula for correction for attenuation due to error of measurement. For the U.S. Employment Service data base, Hunter (Note 4) showed the corresponding "reliability" coefficients to be about .50. Thus, all correlations are low by a factor of .71. The estimate of the correct correlation can be obtained by multiplying the dimension correlation by the reciprocal of .71, which is 1.41. For example, a dimension correlation of .30 based on observed sample validity coefficients corresponds to a correlation of .42 for population validity coefficients.

The dimension correlations are also systematically reduced by error of measurement in applying the job analysis method. This is not relevant to the present application because the job analysis values are to be used in their current imperfect form. However, it is important for theoretical reasons to note that the reliability of the job dimensions varies from one job classification system to another. In particular, the reliability of judgments regarding specific aptitudes is lower than composite judgments for general abilities. The main dimension used to form the job families for the validity generalization analysis (as described in the latter part of this report) is Fine's (1955) worker function dimension of Data. Reanalysis of a reliability study by Cain and Green (1980) shows the reliability of the Data dimension to be .82. Thus, there is a potential increase of 10 percent in the Data-validity correlations if the Data dimension were perfectly measured.

The correct theoretical correlation between a job analysis dimension and a validity coefficient would require correction for both sampling error in the validity coefficient and rater error in the job analysis dimension. The correlation between the Data dimension and the validity of cognitive ability (shown later in Table 11) is .25. If we correct for error of measurement in the Data ratings, this would increase to .28. If we also correct for the effect of sampling error, we obtain a correlation of .39.

Validity Variation Across the Job Spectrum

The variation in validity for each of the three abilities is presented in Table 1 in three ways: the distribution of observed validities, the distribution of validity without sampling error, and the distribution of true validity. Table 1a presents the distribution of observed validity coefficients across the entire job spectrum. Ten percent of the observed validity coefficients lie below .05 for cognitive and perceptual abilities and below .03 for psychomotor ability. Ten percent of the validity coefficients lie above .45 for cognitive and perceptual ability and above .47 for psychomotor ability.

Table 1

Table 1. The Distribution of Observed and True Validity for Three General Abilities (GVN = Cognitive Ability, SPQ = Perceptual Ability, KFM = Psychomotor Ability) Across the Entire Job Spectrum

Table 1a. The Distribution of Observed Validity Coefficients Across All Jobs

	GVN	SPQ	KFM
Mean observed correlation	.25	.25	.25
Uncorrected standard deviation	.15	.15	.17
Observed 10th percentile	.05	.05	.03
Observed 90th percentile	.45	.45	.47

Table 1b. The Distribution of Observed Validity Coefficients Had There Been No Sampling Error, i.e., If All Studies Had Been Done With Samples of 2,000 or More

	GVN	SPQ	KFM
Mean observed correlation	.25	.25	.25
Corrected standard deviation	.08	.07	.11
Corrected 10th percentile	.15	.16	.11
Corrected 90th percentile	.35	.34	.39

Table 1c. The Distribution of True Validity Across All Jobs; i.e., the Distribution of Validity Had Job Performance Been Perfectly Measured and Had the Studies Been Done on Applicant Populations

	GVN	SPQ	KFM
Mean true validity	.47	.38	.35
Standard deviation of true validity	.12	.09	.14
10th percentile of true validity	.31	.26	.17
90th percentile of true validity	.63	.50	.53

Much of the variation in Table 1a is spurious due to the sampling error in the observed validity coefficients. If the effect of sampling error were removed, the variation about the mean would be much smaller. Table 1b presents the distribution of validity coefficients had there been no sampling error. The mean validity is still .25 for each ability, but the variation is substantially reduced, as a result of which the 10th-percentile points are much closer to the means--.15, .16, and .11 for cognitive, perceptual, and psychomotor ability, respectively. These values are all well above zero. Thus, when the validities are corrected for sampling error, we find that none of the three abilities is completely invalid for any job, although psychomotor ability would be essentially zero for a very small number of jobs.

Although the analysis of job classification methods was carried out on observed validity coefficients, it is important to remember that observed validity coefficients are a serious understatement of the validity coefficient that applies to actual job performance. Observed validity coefficients understate true validity for two reasons that we can correct for: error of measurement in the job performance measure (i.e., attenuation due to criterion unreliability) and restriction in range due to using incumbent workers rather than applicants. For comparison purposes, Table 1c presents values computed for the validity generalization analysis to be presented later. Table 1c presents the distributions of true validity; that is, the validity coefficients free of artifacts due to study imperfections. Means are corrected for error of measurement in performance and for range restriction. Variances are corrected for sampling error and for differences between studies in range restriction.

Table 1c shows that only 10 percent of true validity coefficients fall below .31 for cognitive ability and below .26 for perceptual ability. This confirms the conjecture of Schmidt and Hunter (1978) and Hunter (1980) that the major mental aptitudes would be valid for all jobs. Even for psychomotor ability, the mean validity is 2.50 standard deviations above zero. Thus, psychomotor ability will be invalid for fewer than one in 100 jobs. Furthermore, it is worth remembering that we cannot correct for such artifacts as computational and copying errors, criterion deficiency or contamination, and

so forth. Thus, the variation shown in Table 1c is an overestimate by some unknown amount.

Table 1 shows that all three GATB ability composites are valid in predicting performance in all jobs. However, for each of the composites the level of validity will be low for some jobs. A uniformly high level of prediction can only be obtained if job families can be found for which there is ability substitution in prediction.

Test Development Analysts' Judgments

Each of the 515 validation studies done by the U.S. Employment Service began with a job analysis. As part of that job analysis, test development analysts were asked to assess the aptitude requirements of the job by rating each of the nine GATB aptitudes for relevance to that job. These ratings were correlated across jobs and the correlations are presented in Table 2a.

Table 2

Table 2. Basic Results for Test Development Analysts' Judgments; GVN = Cognitive Ability, SPQ = Perceptual Ability, KFM = Psychomotor Ability

Table 2a. Correlations* Between Test Development Analysts' Judgments for Single Aptitudes

		G	V	N	Q	S	P	K	F	M
Intelligence	G	100								
Verbal Aptitude	V	46	100							
Numerical Aptitude	N	45	35	100						
Clerical Perception	Q	31	43	28	100					
Spatial Aptitude	S	38	-2	21	-23	100				
Form Perception	P	-24	-41	-28	-34	11	100			
Motor Coordination	K	-52	-40	-45	-25	-24	24	100		
Finger Dexterity	F	-41	-40	-33	-26	-10	30	37	100	
Manual Dexterity	M	-34	-57	-44	-42	0	30	31	27	100

* Decimals Omitted

Table 2b. Correlations* Between Composite Test Development Analysts' Judgments and Actual Composite Validity

		Actual Validity			TDA Judgment		
		GVN	SPQ	KFM	GVN	SPQ	KFM
Actual Validity	GVN	100					
	SPQ	63	100				
	KFM	6	35	100			
TDA Judgments	GVN	30	1	-42	100		
	SPQ	9	11	-15	25	100	
	KFM	-27	-3	34	-74	-17	100

* Decimals Omitted

The correlations between analysts' judgments follow a pattern which is very different from that reported by Hunter (Note 4) for correlations between aptitudes across people or between actual validities across jobs. The judgments for the three cognitive aptitudes G, V, and N do correlate relatively highly with each other, and the judgments for the three psychomotor aptitudes K, F, and M do correlate relatively highly with each other, but the judgments for the three perceptual aptitudes S, P, and Q do not correlate highly with each other. Instead, analysts see Clerical Perception (Q) as a cognitive aptitude while they see Form Perception (P) as a psychomotor aptitude. Spatial aptitude (S) is not correlated with either. This lack of correlation may result from disagreement between analysts as to the nature of S. Suppose that half the analysts see S as a cognitive aptitude while the other half see S as a psychomotor aptitude. An analysis across all analysts would produce a canceling effect among these disagreements with a resultant pattern of all near-zero correlations--like that in Table 2a.

There is another stark difference between the pattern of correlations between judgments and the pattern of correlation between aptitudes or between aptitude validities. The correlations between the two-judgment clusters are all negative. Indeed, the correlations between judgments about cognitive and psychomotor clusters are as large in magnitude as the positive correlations within each cluster. At the level of factor scores, the analysts' judgments about cognitive aptitudes and the analysts' judgments about psychomotor aptitudes are perfectly negatively correlated. Thus, analysts see only one ability dimension as underlying the job spectrum: a dimension running from mental (high-cognitive, low-psychomotor) jobs at one end to physical (low-cognitive, high-psychomotor) jobs at the other.

Table 2b brings this contrast between actual and perceived validities into sharper focus. Both sets of data are presented as aptitude composites, though the perceptual composite of analyst judgments is actually meaningless.

For actual validities, there is a correlation of +.06 between cognitive and psychomotor composites, whereas for analysts judgments the correlation is -.74. Thus, analysts see only one dimension where there are actually two. This pattern also emerges when we look at the extent to which analysts predict actual validity. The analyst cognitive composite not only positively predicts actual cognitive validity ($\underline{r}$ = .30), but it negatively predicts actual psychomotor validity ($\underline{r}$ = -.42). In fact, the analysts' judgments about cognitive validity actually predict (negatively) psychomotor validity better than they predict cognitive validity. The analyst psychomotor composite positively predicts actual psychomotor validity ($\underline{r}$ = .34) but also negatively predicts cognitive validity ($\underline{r}$ = -.27).

The analyst dimension does not relate in a simple way to either cognitive ability or psychomotor ability alone. Rather, the analysts rate cognitive ability as relevant for jobs on which cognitive ability has high validity <u>and</u> where psychomotor ability has low validity. Analysts rate cognitive ability as irrelevant on jobs for which the validity of cognitive ability is low <u>and</u> psychomotor ability has high validity. Thus, the analyst dimension GVN is like the difference between the validity dimensions: Analyst GVN = Validity GVN - Validity KFM. The perceptual validity dimension is lost to analyst judgments.

<u>DOT Estimated Means</u>

In the construction of the <u>Dictionary of Occupational Titles</u>, the entire work spectrum was subjected to job analysis. As part of this analysis, each of 12,000 jobs was examined for aptitude requirements in terms of the nine GATB aptitudes. Alas, the analysts were not asked to rate each aptitude for relevance to the job directly, but were asked to estimate incumbent means on each aptitude. Incumbent means are determined by market forces as well as by aptitude requirements, and analysts are aware of these market forces. Jobs with high cognitive aptitude requirements tend to pay higher and have higher security. Thus, people with high cognitive ability tend to be concentrated in higher paying jobs. The people left in lower paying jobs tend to have lower cognitive ability, regardless of whether or not the job has high cognitive requirements.

Table 3

Table 3. Basic Results for DOT Estimated Means

Table 3a. Correlations* Between DOT Estimated Means Across Jobs

		G	V	N	S	P	Q	K	F	M
Intelligence	G	100								
Verbal Aptitude	V	83	100							
Numerical Aptitude	N	75	72	100						
Spatial Aptitude	S	56	47	55	100					
Form Perception	P	51	41	45	70	100				
Clerical Perception	Q	53	57	55	19	26	100			
Motor Coordination	K	10	8	5	33	38	2	100		
Finger Dexterity	F	14	13	18	43	50	13	52	100	
Manual Dexterity	M	3	-1	9	48	40	-6	43	60	100

* Decimals Omitted

Table 3b. Correlations* Between Composite DOT Estimated Means and Actual Composite Validity

		Actual Validity			DOT Means		
		GVN	SPQ	KFM	GVN	SPQ	KFM
Actual Validity	GVN	100					
	SPQ	63	100				
	KFM	6	35	100			
DOT Estimated Means	GVN	27	3	-33	100		
	SPQ	24	5	-33	76	100	
	KFM	1	3	-6	11	40	100

* Decimals Omitted

Table 3a presents the correlations across jobs between estimated aptitude means from the DOT. Again the pattern departs sharply from that of correlations between aptitudes across people or from correlations between validities across jobs (Hunter, Note 4). The cognitive and psychomotor clusters are both present, but the perceptual cluster is deviant. Spatial and Form Perception do correlate highly with each other and with both other clusters, but Clerical Perception does not correlate highly with S and P and does not correlate highly with the psychomotor clusters. Thus, the departure from other ures is largely tied to the deviant pattern of Q.

Table 3b presents the correlations between estimated mean composites and actual validity composites. The correlations between mean composites follow the same pattern as that for composite validities (.76, .40, and .11 versus .63, .35, and .06, respectively). In predicting actual validity, the perceptual composite SPQ is redundant with the cognitive composite GVN in positively predicting cognitive validity ($\underline{r}$ = .24 and $\underline{r}$ = .27) while negatively predicting psychomotor validity ($\underline{r}$ = -.33 in both cases). The psychomotor composite does not predict the validity of any aptitude. Thus, the estimated means for psychomotor aptitude do not predict validity, as did the analyst judgments of psychomotor aptitude. This suggests that psychomotor estimated means are more a matter of perceived market conditions than of actual aptitude relevance.

The estimated means for cognitive ability have the same mixed relationship to actual validity as did the analyst judgment for cognitive ability relevance. The estimated means are high for jobs in which cognitive ability has high validity <u>and</u> psychomotor ability has low validity. Estimated means for cognitive ability are low for jobs on which cognitive ability has low validity <u>and</u> psychomotor ability has high validity. Thus, the estimated means for cognitive ability are related to the difference between validity dimensions. The estimated means for psychomotor ability are irrelevant to validity.

<u>Data, People, and Things</u>

Each job in the <u>DOT</u> is rated on three dimensions developed by Fine (1955; Fine & Heinz, 1958) that characterize the level of worker functioning in relationship to data (i.e., information, knowledge, and conceptions), people, and things (i.e., machines, tools, equipment, and products). Fine established categories on each of these dimensions such that the categories would be rank ordered simultaneously on skill and responsibility. In content, the categories on each dimension tend to be ordered by complexity of task.

Table 4

Table 4. Mean Observed Validity for Original and Modified Data Categories; GVN = Cognitive Ability, SPQ = Perceptual Ability, KFM = Psycho-Motor Ability

Table 4a. Mean Observed Validity for Original Data Categories

		Observed Validity			
Name	Number	GVN	SPQ	KFM	Number of Jobs
Synthesizing	0	.33	.20	.11	9
Coordinating	1	.30	.21	.14	52
Analyzing	2	.30	.29	.20	73
Compiling	3	.29	.28	.24	135
Computing	4	.25	.28	.36	16
Copying	5	.21	.22	.25	20
Comparing	6	.21	.23	.31	210

Table 4b. Mean Observed Validity for Modified Data Categories

		Observed Validity			
Name	Number	GVN	SPQ	KFM	Number of Jobs
Synthesize/Coordinate	1	.30	.21	.13	61
Analyze/Compile/Compute	2	.29	.28	.23	224
Copy/Compare	3	.21	.23	.30	230

Table 4 presents average observed validities for the Data categories. Table 4a presents the averages for the original seven categories. Examination of this table shows that there is little information lost by merging categories 2, 3, and 4, or by merging categories 5 and 6. Table 4b shows average validities for the new Data categories. The new categories have two advantages: (1) validity averages for the cognitive composite GVN and the psychomotor composite KFM are now linearly related to category number, and (2) the reduction in the number of categories lends itself to work with other dimensions when the number of combinations grows combinatorially. Table 4b brings out another important fact: The Data categories are related to the perceptual composite SPQ, although in a nonlinear way. The perceptual composite has average validities of .21, .28, and .23 for the new categories 1, 2, and 3, respectively. Thus, SPQ has the highest validity for the middle Data category. If a quadratic trend is scored for Data (i.e., if Data is scored 0-1-0 for categories 1, 2, and 3, respectively), then the validity of SPQ would correlate with the quadratic trend, though not with Data itself.

Table 5

Table 5. Mean Observed Validity for Original and Modified People Categories; GVN = Cognitive Ability, SPQ = Perceptual Ability, KFM = Psycho-Motor Ability

Table 5a. Mean Observed Validity for Original People Categories

		Observed Validity				
Name	Old	GVN	SPQ	KFM	Number of Jobs	Sum
Mentoring	0	.18	.14	.06	10	.34
Negotiating	1	.38	.29	.28	4	.95
Instructing	2	.24	.18	.17	13	.59
Supervising	3	.29	.20	.21	2	.70
Persuading	5	.21	.23	.22	6	.66
Signalling	6	.31	.25	.19	127	.75
Serving	7	.34	.29	.28	15	.91
Helping	8	.23	.25	.28	338	.76

Table 5b. Mean Observed Validity for Modified People Categories

			Observed Validity			
Name	Old	New	GVN	SPQ	KFM	Number of Jobs
Mentoring	0	0	.18	.14	.06	10
Instructing	2	1	.24	.18	.17	13
Supervising/Persuading	3,5	2	.23	.22	.22	8
Signalling/Helping	6,8	3	.25	.25	.26	465
Negotiating/Serving	1,7	4	.35	.29	.28	19

Table 5 presents validity averages for the People job dimension. Table 5a presents the validity averages for the original People categories. It is clear in this table that validity is not consistently ordered for any of the three composites. On the other hand, it would appear that validity follows the same pattern across categories for each of the aptitude composites. Therefore, categories were reordered to reflect this fact. First, the three mean validities were summed across composites for each category. These sums are in the last column in Table 5a. The new categories are old category 0, old category 2, old categories 3 and 5, old categories 6 and 8, and old categories 1 and 7. Old category 4 never occurred and was ignored. The main change in the category system was the shift in the location of old category

1, "negotiation." It seems clear from the frequency and location of this category that it is being used for many jobs in a manner quite different from Fine's original conception. The other change was to invert the order of categories 7 and 8.

Table 5b shows the mean validities for the new People categories. For the new categories, validity is linearly ordered for each of the three ability composites, except that categories 1 and 2 are approximately equal for the cognitive composite GVN.

Table 6

Table 6. Mean Observed Validity for Original and Modified Things Categories; GVN = Cognitive Ability, SPQ = Perceptual Ability, KFM = Psycho-Motor Ability

Table 6a. Mean Observed Validity for Original Things Categories

		Observed Validity			
Name	Old	GVN	SPQ	KFM	Number of Jobs
Set up	0	.34	.35	.19	21
Precision Work	1	.28	.27	.22	139
Operating-controlling	2	.28	.28	.26	89
Driving-operating	3	.23	.19	.20	8
Manipulating	4	.21	.23	.30	85
Tending	5	.22	.24	.30	42
Feeding-offbearing	6	.13	.16	.35	20
Handling	7	.25	.23	.24	111

Table 6b. Mean Observed Validity for Modified Things Categories

			Observed Validity			
Name	Old	New	GVN	SPQ	KFM	Number of Jobs
Set up	0	0	.34	.35	.19	21
Precision work/Driving	1,3	1	.28	.26	.22	147
Controlling/Handling	2,7	2	.27	.25	.25	200
Manipulating/Tending	4,5	3	.22	.23	.30	127
Feeding-Offbearing	6	4	.13	.16	.35	20

Table 6 presents mean validities for the Things dimension. Table 6a presents the averages for the original Things categories. Mean validities are approximately ordered for the cognitive and psychomotor composites, although with certain common discrepancies, most notably for category 7, "handling." The reason for this is that analysts tended to use "handling" for their residual category, that is, for jobs that did not fit into any other Things category. Linearity is slightly improved if categories 2 and 3 are interchanged, that is, if "driving-operating" is placed ahead of "operating-controlling."

Table 6b presents the average validities for the new Things categories. The average validities are perfectly ordered by the new categories for all three composites, although the validities for psychomotor ability are in the opposite order to those for cognitive and perceptual abilities.

Table 7

Table 7. Contingency Tables Relating the Frequencies of Data, People, and Things (Modified Categories)

Table 7a. Data and People

		Data			
		1	2	3	
People	0	10			10
	1	13			13
	2	2	6		8
	3	34	206	225	465
	4	2	12	5	19
		61	224	230	515

Table 7b. Data and Things

		Data 1	Data 2	Data 3	
Things	0	1	19	1	21
	1	26	108	13	147
	2	31	76	93	200
	3	3	21	103	127
	4			20	20
		61	224	230	515

Table 7c. People and Things

		Things 0	Things 1	Things 2	Things 3	Things 4	
People	0		8				10
	1		7	3	3		13
	2	1	2	5			8
	3	20	124	181	120	20	465
	4		6	9	4		19
		21	147	200	127	20	515

The Data, People, and Things categories are not independently used (and may not be independently defined). Table 7 presents three contingency tables showing the frequencies of jobs (from the 515 USES validation studies) in various combinations of new categories. Table 7a presents the breakdown by frequency of validation studies for the new People and Data categories. It is clear from this table that most jobs are placed in People category 3, "signalling/helping." Otherwise, the People dimension is basically a division of Data category 1 in which "mentoring" and "instructing" are separated from the rest. A comparison of the rows for "mentoring" and "instructing" in Table 5b with the row for Data category 1 in Table 4b shows that "mentoring" and "instructing" have the same pattern of validities except that they are uniformly lower than the other jobs in Data category 1. This is largely a matter of restriction in range: "mentoring" and "instructing" characterize jobs held by college graduates. Table 7b shows that Data and Things are highly correlated.

Table 8

Table 8. Correlations* Between Validity and the Data, People, Things Dimensions (Using the Modified Category Numbers)

			GVN	SPQ	KFM	D	T	P
Validity	Cognitive	GVN	100					
	Perceptual	SPQ	63	100				
	Psychomotor	KFM	6	35	100			
Job Dimensions:	Data		25	3	-32	100		
	Things		23	17	-22	50	100	
	People		-10	-10	-18	38	12	100

* Decimals Omitted

The new categories for Data, People, and Things were arranged so that category labels would be linearly related to validity. This suggests that the possibility of combinations of dimensions might be analyzed using linear correlation. The correlations for this analysis are presented in Table 8. This table shows that Data and Thirgs are correlated .50. Thus, the dimension correlations for the Things dimension are in part accounted for by the Data dimension. However, partial correlations (not shown) indicate that the Things dimension would add to the prediction of validity for the cognitive and the perceptual dimensions. The People dimension has a marginally useful level of correlation only for psychomotor ability (r = -.18). However, the People dimension is correlated .38 with the Data dimension, and the partial correlation (not shown) between People and psychomotor validity with Data held constant is only -.07. Thus, the linear analysis suggests that People adds nothing to the prediction made by Data alone while Things would contribute to the prediction of the validity of cognitive ability and to the prediction of perceptual validity as well.

Table 9

Table 9. Mean Observed Validity* for Data-People-Things Combinations; "DPT Code" Digits Stand for the Modified Data, People, and Things Category Numbers Respectively; GVN = Cognitive Ability, SPQ = Perceptual Ability, KFM = Psychomotor Ability

DPT Code	Number of Jobs	Observed Validity GVN	SPQ	KFM	Beta Weight GVN	SPQ	KFM	R_3	Beta Weight GVN	KFM	R_2
101	8	18	16	4	13	8	5	20	14	11	17
102	2	19	4	15	40	-36	20	30	16	10	21
111	7	24	13	16	35	-21	15	28	21	9	25
112	3	23	26	24	9	12	15	30	17	18	29
113	3	22	20	10	16	7	1	22	21	3	22
120	1	35	15	19	58	-38	19	42	32	8	36
122	1	24	26	23	11	11	14	30	18	17	29
131	11	34	17	9	50	-23	3	37	35	-3	34
132	23	36	25	15	41	-9	5	36	35	3	36
142	2	50	41	27	46	1	11	51	46	11	51
221	2	16	16	28	12	-7	28	29	7	26	29
222	4	24	25	20	15	12	11	30	19	13	27
230	19	33	36	18	13	26	0	37	30	7	33
231	100	28	29	23	15	11	12	32	23	15	31
232	69	29	26	22	23	2	13	32	24	14	32
233	18	22	24	33	11	1	29	35	12	29	35
241	6	41	36	34	34	-2	23	46	33	22	46
242	3	28	21	27	31	-14	24	35	21	20	34
243	3	28	19	18	33	-13	13	30	25	9	29
330	1	53	36	23	62	-16	10	54	51	5	53
331	13	20	24	28	6	8	22	30	12	24	30
332	89	22	24	29	11	4	23	32	14	24	32
333	102	21	23	31	11	1	27	33	12	27	33
334	20	13	16	35	5	-6	37	36	1	35	35
342	4	29	26	32	24	-6	27	37	20	25	37
343	1	34	29	3	27	16	14	38	39	-10	36

* Decimals Omitted

If there were an interaction between the job analysis dimensions in the prediction of validity, then the linear analysis of the previous paragraph would overlook important information. A search was made for such interactions. Table 9 presents the mean validity for all combinations of the new categories. Since many cells have only one or two studies, the search for interactions requires combining cells. Many such combinations were tried and none showed evidence of interaction.

Table 10

Table 10. Mean Observed Validity* for the Categories of Job Complexity Created Using the Modified Data and Things Categories and the Implied Beta Weights for Ability Combinations; GVN = Cognitive Ability, SPQ = Perceptual Ability, KFM = Psychomotor Ability

		Validities			Beta-Weights				
Complexity Levels		GVN	SPQ	KFM	GVN	SPQ	KFM	R	Number of Jobs
Setup	1	34	35	19	18	20	3	37	21
Synthesize/Coordinate	2	30	21	13	34	-7	5	31	60
Analyze/Compile/Compute	3	28	27	24	21	3	15	32	205
Copy/Compare	4	22	24	30	9	5	25	33	209
Feeding/Offbearing	5	13	15	35	5	-6	37	36	20

* Decimals Omitted

After the search for interactions was abandoned, a search was made for a category scheme to capture the linear analysis. After examining various Data-Things combinations, it became clear that the contribution of the Things dimension is its extreme categories: industrial setup work and feeding-offbearing jobs. Industrial setup work is extremely complex while feeding and offbearing jobs are among the simplest jobs in the economy. If these categories are added to the collapsed three Data categories, then a system of five job families emerges. These five categories are shown in Table 10. This extended version of the Data dimension will be called "job complexity."

The pattern of mean validity across levels of job complexity is that predicted by the correlational analysis. As complexity decreases, the mean validity of cognitive ability decreases from .34 to .13 and the mean validity of psychomotor ability increases from .19 to .35. The validity of perceptual ability is strikingly high for industrial setup work.

In addition to computing the mean validity for each ability considered separately, it is possible to compute beta weights for the abilities considered jointly. These beta weights are also shown in Table 10. For the bottom four categories, the results are what would be predicted from the consideration of validity for single abilities. The beta weight for cognitive ability drops as complexity decreases while the beta weight for psychomotor ability rises. Perceptual ability is redundant for these four levels of complexity and makes no contribution to the multiple regression.

The results for industrial setup work are striking in that perceptual ability plays a key role in the multiple regression. The dependence of setup work on perceptual ability is the main reason that the Things dimension correlates with perceptual ability.

Job complexity basically reflects the Data dimension. Thus, again we have one dimension in the job analysis scheme predicting two dimensions of validity. This can be seen in Table 8, which showed that Data correlates .25 with cognitive validity but -.32 with psychomotor validity. It can be seen in Table 10 in the form of the high negative correlation between cognitive and psychomotor validity across categories, a correlation of -.94. As with the relevant dimensions in the first two job classification systems, the Data dimension relates to the difference between validity for cognitive and psychomotor ability. That is, the Data dimension is high when cognitive ability is highly valid <u>and</u> psychomotor validity is low, while Data is low when cognitive validity is low <u>and</u> psychomotor validity is high.

Table 11

Table 11. Correlations Between the Dimensions of Three Job Analysis Systems; GVN = Cognitive Ability, SPQ = Perceptual Ability, KFM = Psycho-Motor Ability

<u>Table 11a</u>. Correlations* Between Observed Validity Coefficients, Test Development Analysts' Judgments, <u>DOT</u> Estimated Means, and Modified Data-People-Things Codes

		Validity			TDA			DOT EM			D-T-P		
		GVN	SPQ	KFM	GVN	SPQ	KFM	GVN	SPQ	KFM	DAT	THI	PEO
Validity	GVN	100											
	SPQ	63	100										
	KFM	6	35	100									
Test Development Analyst	GVN	30	1	-42	100								
	SPQ	9	11	-15	25	100							
	KFM	-27	-3	34	-74	-17	100						
<u>DOT</u> Estimated Means	GVN	27	3	-33	75	30	-66	100					
	SPQ	24	5	-33	55	41	-42	76	100				
	KFM	1	3	-6	-4	15	20	11	40	100			
D-T-P	Data	25	3	-32	71	27	-64	85	61	3	100		
	Things	23	17	-22	41	40	-29	51	53	27	50	100	
	People	-10	-15	-18	24	-7	-27	41	27	14	38	12	100

* Decimals Omitted

Table 11b. Subset of Table 11a Showing the Redundancy Among the Relevant Job Dimensions From Different Systems*

		Validity			TDA	TDA	DOT	
		GVN	SPQ	KFM	GVN	-KFM	GVN	Data
TDA	GVN	30	1	-42	100	74	75	71
TDA	-KFM	27	3	-34	74	100	66	64
DOT	GVN	27	3	-33	75	66	100	85
	Data	25	3	-32	71	64	85	100

* Decimals Omitted

Redundancy of the Dimensions Relevant to Validity

Table 11 presents the correlations between all the job analysis dimensions so far defined, and the correlations with actual validity as well. Table 11a presents the entire correlation matrix, while Table 11b presents the key entries in Table 11a. Table 11b brings out the redundancy in the dimensions of validity; that is, the very high correlations between the test development analyst judgments of GVN and KFM (reverse scored in Table 11b), the DOT estimated means for GVN, and the Data dimension. The right block in Table 11b shows these dimensions to be very highly correlated with each other. The left block shows that they all have virtually identical correlations with actual validity.

Thus, all of the information in the job analysis methods considered to this point is captured in the job complexity categories defined by the extended version of the Data dimension in Table 10.

The Position Analysis Questionnaire (PAQ): Study 1

The Position Analysis Questionnaire (McCormick, Jeanneret, & Mecham, 1972) is a set of 189 items pertaining to the type of information processes used by the worker, the job processes carried out, and the working conditions and responsibilities of the job. Two studies have been done relating the dimensions of the PAQ to the validity of cognitive, perceptual, and psychomotor ability: McCormick, Jeanneret, and Mecham (1972) and Mecham, McCormick, and Jeanneret (1977). This section will review the first study. The upshot is that the PAQ appears to have only one relevant dimension and that dimension appears to be job complexity. The second study will be reviewed in the next section. Those results point in the same direction.

McCormick, Jeanneret, and Mecham (1972) had PAQ data on 179 positions and the results of 90 GATB validation studies done by the U.S. Employment Service. They split their position sample into random halves for cross-validation

purposes, and then did a multiple regression of each of the nine GATB aptitude validities onto the job dimension scores on the PAQ for each position. The effective sample size for each multiple regression was 90, with nine predictors. There was substantial capitalization on chance, with the result that on cross-validation their multiple correlations fell from an average of .46 to an average of .24. However, the cross-validated regression equations did predict aptitude validity with r = .36 for the cognitive aptitudes, r = .23 for the perceptual aptitudes and r = .21 for the psychomotor aptitudes. The size of these correlations suggests that the PAQ was tapping into the same job complexity dimension as that tapped by the other job analysis strategies considered herein (the r = .23 for perceptual aptitudes is comparable to the r = .18 for the quadratic Data indicator). However, insufficient specific data are given in their research reports (including various technical reports as well as their journal articles) to adequately test this hypothesis.

For each aptitude, McCormick et al. (1972) list the nine dimensions which correlate most highly with validity for that aptitude. Such a list is highly subject to chance variation. However, by taking the three aptitudes in each ability cluster as replications of one another, this chance variation can be reduced somewhat.

The lists for the cognitive aptitudes on the job-oriented dimensions had four common dimensions: JA-8, 9, 14, and 26, or decision-making, information processing, handling and manipulating activities, and structured work (which probably has a negative beta weight). The lists for the cognitive aptitudes on the attribute-oriented dimensions also had four common dimensions: AA-8, 18, 19, and 21, or information processing, unstructured responsible work, paced structured work (which probably has a negative beta weight), and merit income. According to the coordinated dimension list in McCormick et al.'s Table 6, the cognitive aptitudes are mainly predicted by two dimensions: information processing and unstructured work. This is essentially the same as the Data dimension of the present study.

The lists for the perceptual aptitudes on the job-oriented dimensions had three common dimensions: JA-1, 9, and 17, or visual input, decision-making, and communication of decisions. The perceptual aptitude lists for the attribute dimensions had three common dimensions: AA-6, 7, and 12, or verbal-auditory input, use of job-related knowledge, and interpersonal communication. The only common dimension between the two sets is that of interpersonal communication. This is essentially the same as the People dimension of the present study, and the degree of prediction of perceptual aptitude validity by the PAQ (r = .23) is not far from that found in this study for the People dimension (r = .15).

The lists for the psychomotor aptitudes on the job-oriented dimension had no common dimensions. The lists for the attribute-oriented dimensions had three common dimensions: AA-4, 6, and 12, or non-visual input, verbal-auditory

input, and interpersonal communications. Two of these dimensions are also the common dimensions for the perceptual aptitudes. It is hard to interpret these results. Perhaps the lack of consistency accounts for the fact that the level of prediction of psychomotor aptitude validity by the PAQ (r - .21) is so much less than that found in this study for the Data dimension (r - .32).

Based on this analysis, it appears that the PAQ multiple regression equations tap the same job complexity dimension used by other job analysis systems.

The PAQ: Study 2

Mecham, McCormick, and Jeanneret (1977) revised the PAQ and recomputed correlations between job dimensions and aptitude validity across 163 jobs. From these correlations it is possible to compute the correlations for the three-ability composite validities. These correlations are presented in Table 12.

Table 12

Table 12. The Correlations* Between Validity Coefficients and the Overall Dimensions of the Position Analysis Questionnaire (Computed from Mecham, McCormick, and Jeanneret, 1977, p. 128); GVN = Cognitive Ability, SPQ = Perceptual Ability, KFM = Psychomotor Ability; the Number of Jobs is 163.

Job Dimension	GVN	SPQ	KFM
Decision making/communication	20	-2	-30
Machine operator	-1	19	-2
Clerical	18	-3	-9
Technical	23	4	-21
Service	-17	-15	9
Regular day schedule	-3	-3	5
Routine/repetitive	-22	-2	12
Environmental awareness	2	-1	1
Physical activity	-19	-8	7
Supervising	6	-1	5
Public contact	-9	-6	1
Hazardous	0	3	1
Special schedule or apparel	-17	-23	-7
Combined Dimensions	**GVN**	**SPQ**	**KFM**
Mental work	30	1	-36
Menial work	-38	-11	19
Overall	48	8	-39

* Decimals Omitted

The job dimensions on the PAQ were obtained by use of exploratory factor analysis and are thus artificially orthogonalized. The dimensions as computed are thus not conceptually independent, but may jointly measure a conceptual dimension. The dimension most similar in content to the present study's Data dimension is the PAQ decision-making dimension. The correlations between each of these dimensions and the three ability composite validities (cognitive, perceptual, and psychomotor) are quite similar: .20, -.02, and -.30 respectively (see Table 12) for decision making, in comparison to .25, .03, and -.32 (see Table 11) for Data. The PAQ technical dimension is most similar in content to this study's Things dimension. The correlations again are quite similar: .23, .04, and -.21 (see Table 12) for technical, versus .23, .17, and -.22 (see Table 11) for Things. The combination of these dimensions should therefore be similar to the job complexity dimension. This combined dimension is shown in Table 12 as "mental work" and has correlations of .30, .01, and -.36 in comparison to .29, .12, and -.30 for the job complexity categories in Table 10.

Four other PAQ dimensions correlate with validity in a similar manner: clerical (reverse scored), service, routine, and physical activity. The combination of these four dimensions is referred to as "menial work" in Table 12 and shows correlations with the cognitive, perceptual, and psychomotor composite validities of -.38, -.11, and .19, which is essentially the inverse of the pattern of correlations for the "mental work" index. Menial work was combined subtractively with mental work to form the "overall" dimension shown in Table 12. The validity correlations for this overall dimension are .48, .08, and -.39, in comparison with .29, .12, and -.30 for the job complexity categories of Table 10.

Although there are 13 dimensions to the PAQ, these dimensions are only indirectly relevant to validity. When the dimensions are combined in an optimal way, the working dimension appears to be the same complexity dimension as in other job analysis systems. The overall dimension correlates positively with cognitive validity but negatively with psychomotor validity. That is, when the overall dimension is high for a job, then cognitive ability is highly valid and psychomotor ability has low validity. If a job is low on the overall dimension, then psychomotor ability has high validity and cognitive ability has low validity. Thus, the working dimension that can be derived from the PAQ also appears to predict the difference between cognitive and psychomotor validity, rather than either validity separately.

The fact that the correlations for the PAQ overall dimension were higher than the correlations for the job complexity dimension defined from this study's Data and Things dimensions is in part due to a difference in samples. The correlations between validity coefficients show that the sample of jobs used by Mecham, McCormick, and Jeanneret (1977) was more heterogeneous in complexity than the jobs sampled in the validation studies used in this study. The PAQ correlations are thus somewhat higher because of enhancement of range. However, the data necessary to assess the full extent of the

difference in range are not available. It may be that the PAQ overall dimension measures job complexity better than does the Data and Things composite.

The OAP Structure

The previous job analysis schemes were analytical in nature: They defined job categories in terms of ordinal dimensions. The effective dimension in each system is job complexity. The current Occupational Analysis Patterns structure is an interest-oriented set of job families, though aptitude was considered in the later developmental stages and may have introduced the same job complexity dimension at that point.

The OAP structure began with the development of job families based on occupational interest, presented in the Guide for Occupational Exploration (U.S. Department of Labor, 1979). Eleven interest areas were broken into 66 specific groups of jobs, referred to hereafter as the GOE groups. However, when Droege and Boese (Note 1) sought to use the GOE groups for purposes of predicting aptitude patterns, many revisions were made. First, seven groups were believed to be too heterogeneous to be worth refining and were dropped. Thus, the OAP structure to be considered here is not exhaustive. Next, seven groups were broken down on the basis of the Data code to form 14 groups. This produced a new structure of 66 modified GOE groups. These 66 groups were then reduced to 22 OAP groups by a complicated two-stage regression analysis in which the results of the validation studies were used. First, those groups with more than one validation study were selected for special consideration. Within each such group, the multiple-cutoff patterns generated by the validation studies in that group were compared and a "modal aptitude pattern" was defined. Nine criterion variables were then formed for each validation study based on whether or not each aptitude was in the modal pattern. For each of the nine criterion variables, a nine-variable multiple regression equation was developed using the DOT estimated means as predictors. These regression equations were then used to collapse the groups. Certain other changes were alluded to by Droege and Boese (Note 1), but not described in detail.

Table 13

Table 13. Average Observed Validity* for the OAP Job Categories; GVN = Cognitive Ability, SPQ = Perceptual Ability, KFM = Psychomotor Ability

		Observed Validity			Beta Weights			
Group	Number of Jobs	GVN	SPQ	KFM	GVN	SPQ	KFM	R
2	3	19	17	16	16	0	11	22
3	14	27	16	18	35	-18	15	30
5	131	27	29	24	13	12	13	32
6	16	27	19	7	30	-2	-2	28
7	7	29	22	9	32	1	-1	31
9	15	21	22	32	13	-2	29	34
10	22	28	21	14	29	-4	6	28
11	13	36	18	10	53	-24	4	39
12	7	32	29	19	24	10	7	35
13	19	25	26	23	14	9	14	30
15	95	24	26	28	12	7	21	33
16	124	19	22	34	8	0	31	35
17	10	32	23	13	35	-5	4	33
18	22	31	25	19	29	-2	10	32
19	2	34	25	19	36	-8	10	35
22	15	37	31	22	33	1	7	37

* Decimals Omitted

Table 13 shows the average validities for the resulting 22 OAP groups and the beta weights derived from them. Data were available for 16 of the 22 groups. The multiple regression equations break the groups into three categories: First, eight of the groups end up with prediction based solely on the cognitive composite, GVN; these are groups 3, 6, 7, 10, 11, 17, 18, 19, and 22, representing 114 of the 515 validation studies. Second, three of the groups end up with prediction based almost solely on the psychomotor composite KFM; these are groups 9, 15, and 16, representing 234 of the 515 validation studies. Third, four groups have mixed regression equations; these are groups 2, 5, 12, and 13, representing 167 of the 515 validation studies. Even in the mixed-regression groups, the multiple regression does not improve a great deal over the best single predictor (the cognitive composite for groups 2 and 12 and the psychomotor composite for groups 5 and 13).

Table 14

Table 14. The Relationship Between the OAP Categories and the Data Dimension

OAP Category	Data 1	Data 2	Data 3	
Use GVN	57	47	10	114
Mixed	4	154	9	167
Use KFM		23	211	234
	61	224	230	515

Table 14 shows the relationship between the three OAP categories and the Data dimension. Almost all jobs in Data category 3 ("copy/compare") are coded into OAP groups which lead to the use of the psychomotor composite for prediction. Almost all of the jobs in Data category 1 ("synthesize/coordinate") are coded into OAP groups which lead to the use of the cognitive composite for prediction. However, the jobs in Data category 2 ("analyze/compile/compute") do not all go into the OAP mixed regression category; a sizable minority go into the cognitive groups.

Table 15

Table 15. The OAP Groups in the Category "Use GVN" and Their Breakdown by the Data Dimension

OAP Groups	Data 1	Data 2	Data 3
6	16		
11	11	2	
18	18	4	
3	7	7	
10	2	18	
19		2	
22	3	10	2
17		4	6
6+11+18	45	6	
3+10+19+22+17	12	41	10

Table 15 presents the OAP groups in the cognitive category and their relationship to the Data dimension. The jobs in the first three groups (6, 11, and 18) are almost entirely in Data category 1, but the jobs in the last five groups (3, 10, 19, 22, and 17) are more often in Data category 2, and group 17 has a majority of jobs in Data category 3. This latter set of five groups probably represents a clarification of the Data structure generated by the OAP structure. The feeding-offbearing category was lost in the OAP structure; all 20 jobs were buried in group 16. Six of the seven jobs in group 12 were from the industrial setup category, but the other 15 jobs were mostly buried in group 5.

Critique of the OAP Structure

Does the OAP structure capitalize on chance in fitting the present validity data? The modal aptitudes were developed on 430 jobs each of which contributed two pieces of information (since aptitude validity is two-dimensional). Each multiple regression equation developed 10 constants for 90 parameters in all. It would appear that there was considerable room for capitalization on chance, but the two-stage structure of the computational procedure makes it difficult to estimate the exact extent to which this may have occurred.

The second criticism of the OAP structure is related to the first. It is difficult enough to try to capture the conceptual scheme which was worked out in the holistic judgment process in forming the GOE groups. But there is no conceptual process in the multiple regression techniques. Thus, it is hard to know how to check the system for error, and hence hard to know how to improve it.

If the identification of additional job categories which have purely cognitive prediction is a real improvement on the Data-People-Things system (as opposed to capitalization on chance), and this may well be the case, then we need to figure out what new information is being added. That new information might provide a clue to the missing dimension in job analysis.

Job Classification Systems: General Conclusions

Five approaches to job analysis have been evaluated here for their power in predicting aptitude validity. All function about equally well; they predict observed validity to a level of $r = .30$. Corrected for sampling error (as in Hunter, Note 3), this figure increases to about .45. This is quite large enough for practical work, and will provide a very substantial improvement over considering all jobs together. In fact, the improvement will be about half of that which would be realized by a perfect system. It is especially noteworthy that the job complexity dimension which is tapped by these job analyses tends to focus on jobs where either cognitive ability has high validity and psychomotor ability has low validity or where psychomotor

ability has high validity and cognitive ability has low validity. That is, the job complexity dimension tends to locate those jobs which profit from ability substitution for prediction.

Since the Data-People-Things codes are available for all the jobs in the Dictionary of Occupational Titles, the job complexity categories of Table 10 will be carried forward to a full validity generalization analysis.

All of the methods of job analysis considered here reach the same upper bound in capacity to predict aptitude validity. This is because all tap the same relevant dimension, although in very different ways. Aptitude validity is two-dimensional, and hence there must be a missing dimension in the present systems of job analysis. Improvement over the five job families defined by job complexity can only come from finding a job content dimension that corresponds to the missing dimension.

VALIDITY GENERALIZATION

Artifact Distributions

Imperfections in the way that we do validation studies introduce errors into the final results. Validity generalization eliminates certain of these errors: sampling error, error in measurement of job performance, and the artifact of restriction in range. Since the GATB is to be used without improvement in the near future, the present analysis will not correct for error of measurement in ability.

There are usually only a small number of workers who perform the same job in any given work setting. This sets a limit on the sample size in a validation study. When results from small-sample studies are reported, they are subject to large random errors and hence show considerable variation from study to study. The mean correlation across studies varies in relation to the total sample size across the studies. The variance of correlations across studies can be corrected, using a known formula for sampling error, by subtracting the variance due to sampling error.

Table 16

Table 16. The Distribution of People and Jobs Across Levels of Job Complexity (USES Data Base) in Relationship to the Distribution of People in the National Workforce

		Job Proficiency		Training Success		
Complexity Level		Number of Jobs	Number of Persons	Number of Jobs	Number of Persons	Percent of Workforce
Setup	1	17	1,114	4	235	2.5
Synthesize/ Coordinate	2	36	2,455	24	1,863	14.7
Analyze/Compile/ Compute	3	151	12,933	54	3,823	62.7
Compare/Copy	4	201	14,403	8	575	17.7
Feeding/Offbearing	5	20	1,219	0	0	2.4
Total		425	32,124	90	6,496	100.0

Table 16 shows the distribution of jobs and studies for the U.S. Employment Service data base. Of the 515 validation studies, 425 predicted a measure of job proficiency or performance, while 90 predicted training success. There were a total of 38,620 workers who participated in these studies, resulting in an average sample size of 75. Total sample size is over 1,000 for each complexity level for proficiency studies, but is under 1,000 for three of the five levels for training success studies. There were no training success studies for feeding-offbearing jobs. The mean correlations for training success in setup jobs are subject to noticeable sampling error since the total sample size is only 235.

The distribution of people and jobs in the validation studies can be compared to the distribution of people in the national workforce. The workforce distribution is given in the last column of Table 16. More than half the workforce is employed in jobs in complexity level 3. The job proficiency studies are less complex on the average, with a very large overrepresentation of jobs at complexity level 4. This causes distortion in any average across all jobs since average values for proficiency studies will be closer to lower complexity values than would be true for a perfectly stratified sample. The training success studies are closer to the national distribution, although there is an underrepresentation of jobs of lower complexity.

Variation in the reliability of job performance measures stems largely from the difference between studies of training success and studies of job proficiency. Training success is usually assessed with a job knowledge test,

which is likely to have a relatively high reliability (we will assume it to be .80 for the present analysis). Job proficiency is usually measured by the ratings of a single supervisor. King, Hunter, and Schmidt (1980) have shown in a cumulative study that the correct interrater reliability of such a measure is .60 (or less, to the extent that the supervisor's judgment is not perfectly measured). No correction for variation in criterion reliability within study types was made in this study.

The extent of restriction in range in each study can be computed in the U.S. Employment Service studies. First, the standard deviation of each aptitude was recorded for each validation study. This is the incumbent standard deviation. Second, by pooling the data across studies, the total standard deviation for the work population as a whole can be estimated. That is, the mean and standard deviation for each validation study are recorded. A basic formula from analysis of variance shows that the total variance is the variance of the means plus the mean of the variances. We can thus compute the mean and variance of the entire population from the means and variances of each study. This was done for the 425 proficiency studies and produced standard deviations of 61.63, 50.28, and 48.89 for cognitive, perceptual, and psychomotor ability, respectively. These are the applicant standard deviations for each study. The standard measure of restriction in range is the ratio of the incumbent standard deviation to the applicant standard deviation. This parameter (*u*) is 1 if there is no restriction in range and it is lower depending on the extent of restriction.

Table 17

Table 17. The Distribution of Range Restriction Across Levels of Job Complexity; GVN = Cognitive Ability, SPQ = Perceptual Ability, KFM = Psychomotor Ability; Parameter *u* is Defined as the Ratio of the Incumbent Standard Deviation to the Applicant Standard Deviation

Table 17a. The Mean Value of *u*

	Proficiency			Training Success		
Complexity Level	GVN	SPQ	KFM	GVN	SPQ	KFM
1	.67	.81	.89	.64	.73	.89
2	.63	.80	.92	.55	.76	.84
3	.66	.81	.91	.60	.76	.90
4	.68	.83	.89	.66	.83	.94
5	.71	.87	.93	-	-	-
Average	.67	.82	.90	.60	.76	.89

Table 17b. The Standard Deviation of *u*

	Proficiency			Training Success		
Complexity Level	GVN	SPQ	KFM	GVN	SPQ	KFM
1	.053	.065	.074	.093	.049	.026
2	.079	.080	.072	.054	.068	.071
3	.088	.078	.082	.078	.076	.080
4	.082	.089	.083	.066	.112	.130
5	.098	.076	.087	-	-	-
Average	.083	.083	.082	.071	.076	.080

Table 17 shows the distribution of restriction in range in the U.S. Employment Service studies. Table 17a shows the mean value of the range restriction parameter *u* for each ability and for each level of job complexity in both the proficiency and training success studies. The average values across all jobs are .66, .81, and .90 for cognitive, perceptual, and psychomotor ability, respectively. There is much less restriction in range for perceptual and psychomotor ability than for cognitive ability. In fact, these values are such as to suggest that there is little direct selection for perceptual and psychomotor ability. If there were only indirect selection due to the selection on cognitive ability, then the values of *u* for perceptual ability and psychomotor ability would be .82 and .96, in comparison to the observed values of .81 and .90. There is little variation in the average value of *u* across complexity levels.

Table 17b shows the standard deviation of *u* for each ability, for each complexity level, and for proficiency and training success. There is only slight variation in this value across complexity levels or between training and proficiency studies. The standard deviation of restriction in range is about 10 percent of the mean value. The values in Table 17b were used to correct validity standard deviations for differences between studies in range restriction.

Given the distributions of sampling error, error of measurement in job performance, and restriction of range it is possible to transform the distributions of observed validity coefficients into distributions of true validity coefficients. However, this transformation is not perfect since it makes no provision for other artifacts. For example, even with the double-checking that takes place at each level of Employment Service data gathering, there are still recording errors, computational errors, and transcriptional errors. In addition, we know that all criterion measures are likely to be contaminated or deficient to some unknown extent. This means that the variance of true validity is overstated by some unknown amount in this report.

Average Validity

This section will present the findings for average true validity. The following section will present the findings for variation in validity. This section will first present the findings for validity of abilities considered individually. This will be followed by an analysis of validity when prediction is based on all abilities considered together.

Table 18

Table 1. Average True Validity; GVN = Cognitive Ability, SPQ = Perceptual Ability, KFM = Psychomotor Ability

Table 18a. Average True Validity Across the Job Spectrum for Job Proficiency and Training Success Separately

Study Type	Number of Jobs	GVN	SPQ	KFM	Average
Training Success	90	.54	.41	.26	.40
Job Proficiency	425	.45	.37	.37	.40
Average	515	.47	.38	.35	.40

Table 18b. Average True Validity as a Function of Job Complexity

Complexity Level	Proficiency			Training Success		
	GVN	SPQ	KFM	GVN	SPQ	KFM
1	.56	.52	.30	.65	.53	.09
2	.58	.35	.21	.50	.26	.13
3	.51	.40	.32	.57	.44	.31
4	.40	.35	.43	.54	.53	.40
5	.23	.24	.48	-	-	-
Average	.45	.37	.37	.55	.41	.26

Table 18 presents the findings for average true validity. Table 18a shows the average validity across all jobs, with separate subaverages for proficiency and training success. The average validity across all abilities is .40 for both training success and job proficiency. However, the averages

are not the same for individual abilities. Cognitive ability predicts training success (r = .54) slightly better than it does job proficiency (r = .45). Psychomotor ability predicts job proficiency (r = .37) much better than it does training success (r = .26).

Table 18b shows the average validity for each level of job complexity. The mean validity coefficients for job complexity show the same pattern as the average raw correlations: The validity of cognitive ability drops from .56 to .23 as job complexity decreases while the validity of psychomotor ability increases from .30 to .48. That is, cognitive validity is positively related to job complexity. Psychomotor ability has a high validity for predicting job performance in complexity levels 4 and 5, where the validity of cognitive ability is lowest.

The results for training success shown in Table 18b differ from the results for job proficiency. The validity of cognitive ability for predicting training success is uniformly high across levels of job complexity. This result would be expected from findings in learning studies showing that general cognitive ability is predictive of learning in all contexts. The validity of psychomotor ability varies even more sharply for training success than for job proficiency, from .09 to .40 for the four categories with data (as opposed to .21 to .48 across all five categories for job proficiency).

The finding that the validity of cognitive ability is high for all levels of job complexity has been subsequently replicated in an analysis of the U.S. Navy validation data base. Pearlman (1982) applied many job classification systems, including the present complexity dimension, to 500 validation studies (with a criterion of training success) in 61 enlisted Navy occupations. He found no classification scheme which produced more than minimal differences in the validity of cognitive ability for predicting training success. His finding of an average cognitive test validity of .56 is nearly identical to that found in the U.S. Employment Service data base.

Cognitive ability predicts job proficiency for all jobs to a useful extent. However, the validity drops off sharply for low levels of job complexity. These are exactly the same jobs for which psychomotor ability has its highest validity. Thus, validity can be substantially improved by considering more than one ability. The average validity for various combinations of abilities is shown in Table 19.

Table 19

Table 19. Validity of Ability Combinations; GVN = Cognitive Ability, SPQ = Perceptual Ability, KFM = Psychomotor Ability

Table 19a. Validity of Ability Combinations for Job Proficiency: Best Single Predictor, and Two Sets of Multiple Regression Weights With Multiple Correlation

Complexity Level	Best Single Predictor	Beta Weights GVN	SPQ	KFM	R_3	Beta Weights GVN	KFM	R_2
1	.56	.40	.19	.07	.59	.52	.12	.57
2	.58	.75	-.26	.08	.60	.58	.01	.58
3	.51	.50	-.08	.18	.53	.45	.16	.53
4	.43	.35	-.10	.36	.51	.28	.33	.50
5	.48	.16	-.13	.49	.49	.07	.46	.49
Average	.48	.42	-.09	.27	.51	.37	.24	.52

Table 19b. Validity of Ability Combinations for Training Success: Best Single Predictor, and Two Sets of Multiple Regression Weights With Multiple Correlation

Complexity Level	Best Single Predictor	Beta Weights GVN	SPQ	KFM	R_3	Beta Weights GVN	KFM	R_2
1	.65	.57	.21	-.21	.68	.70	-.16	.66
2	.50	.72	-.30	.03	.53	.52	-.05	.50
3	.57	.57	-.07	.15	.58	.53	.13	.59
4	.54	.34	.17	.20	.59	.46	.24	.59
5	-	-	-	-	-	-	-	-
Average	.55	.59	-.10	.11	.57	.53	.08	.57

Table 19a shows the findings for the validity of ability combinations for the prediction of job proficiency. The first column of Table 19a shows the validity of the best single predictor for each level of job complexity. This is the validity of cognitive ability for levels 1, 2, and 3, and psychomotor ability for levels 4 and 5. If only cognitive ability were used, the validity could be as low as .23 at complexity level 5. If only psychomotor ability were used, then the validity could be as low as .21 at complexity level 2. However, if the best single predictor is used at each level, then the lowest validity is .43 at complexity level 4.

Abilities can also be combined using multiple regression. Two sets of multiple regression results are shown in Table 19a. The middle four columns present the multiple regression for all three abilities while the last three columns present the multiple regression results with perceptual ability left out. Of the middle four columns, the first three are the beta weights for the three abilities and the fourth column contains the corresponding multiple correlation. For setup work, all three beta weights are positive and the multiple correlation is .59, which represents a small improvement over the validity of .56 for cognitive ability alone.

For the other four complexity levels, the beta weight for perceptual ability is negative. Since it is unlikely that there are true suppressor effects for perceptual ability, these negative beta weights can be regarded as an artifact of the imperfect measurement of the three abilities. That is, Hunter (Note 4) found that perceptual ability would be superfluous for most jobs if all three abilities were perfectly measured. The beta weight for perfectly measured variables would be zero for such variables.

The last three columns of Table 19a present the regression analysis for cognitive and psychomotor ability. These beta weights track job complexity even more sharply than do the simple validity coefficients. The beta weight for cognitive ability varies from .58 down to .07 while the beta weight for psychomotor ability varies from .01 to .46. The beta weights, however, would suggest that setup work is not higher in complexity than the synthesize/coordinate category, but rather just below it. Both beta weights fall into place if the interchange is made, with .52 between .58 and .45 while .12 is between .01 and .16.

Table 19b shows the results of ability combinations for the prediction of training success. At all levels of complexity, cognitive ability is the best predictor. Thus, the results for the "best single predictor" are identical to the average validity for cognitive ability. The regression equations for all three abilities show negative weights for three of four complexity levels. The negative weight for setup work may reflect sampling error; the total sample size for that category is only 235. There are also negative beta weights for psychomotor ability for complexity levels 1 and 2 when perceptual ability is left out, although they are trivial in magnitude. For complexity levels 3 and 4, the multiple correlations are just as high for cognitive and psychomotor ability as for all three abilities. In fact, psychomotor ability only adds to the multiple correlation for complexity level 4. Overall, there is only a trivial improvement in prediction made by using either perceptual or psychomotor ability in addition to cognitive ability.

Table 20

Table 20. Recommended Regression Equations for Predicting Job Performance and Training Success at Each Level of Job Complexity; EJP = Expected Job Performance, ETS = Expected Training Success; GVN = Cognitive Ability, SPQ = Perceptual Ability, KFM = Psychomotor Ability

Table 20a. Recommended Regression Equations for Predicting Job Proficiency

Complexity Level	Regression Equation	Multiple Correlation
1	EJP = .40 GVN + .19 SPQ + .07 KFM	.59
2	EJP = .58 GVN	.58
3	EJP = .45 GVN + .16 KFM	.53
4	EJP = .28 GVN + .33 KFM	.50
5	EJP = .07 GVN + .46 KFM	.49

Table 20b. Recommended Regression Equations for Predicting Training Success

Complexity Level	Regression Equation	Multiple Correlation
1	ETS = .65 GVN	.65
2	ETS = .50 GVN	.50
3	ETS = .53 GVN + .13 KFM	.59
4	ETS = .46 GVN + .24 KFM	.59
5	-	-

The regression equations recommended for operational use of the current results are presented in Table 20. For job proficiency, the average multiple correlation is .53. This is a substantial improvement over the average of .48 for the best single predictor, which was a substantial improvement over the average of .45 for cognitive ability alone. For training success, the average multiple correlation is .57, which is an improvement over an average of .55 for cognitive ability alone.

Variation in Validity: Homogeneous Hiring

Validity varies from job to job. Part of that variation is accounted for by the job complexity dimension, but there is also variation within the complexity categories due to other unknown dimensions. This variation can be measured and the results for the U.S. Employment Service data base will be

presented in this section. The practical import of such variation is that if an employer bases a selection program on the Employment Service data base, then the actual validity will vary by some random amount from the mean values of Table 20. The extent of such variation depends on the nature of hiring. We will define hiring as "homogeneous" if all applicants are to be hired for exactly the same job. Hiring will be called "heterogeneous" if different applicants are to be hired for different jobs. The extent of random departure from the mean values of Table 20 is maximal for homogeneous hiring. The reduction in random variation under heterogeneous hiring will be discussed in the next section.

Table 21

Table 21. Variation in Validity Within Job Complexity Categories; GVN = Cognitive Ability, SPQ = Perceptual Ability, KFM = Psychomotor Ability

Table 21a. The Standard Deviation of True Validity

	Proficiency			Training Success		
Complexity Level	GVN	SPQ	KFM	GVN	SPQ	KFM
1	.03	.00	.04	.00	.16	.00
2	.15	.00	.00	.20	.00	.09
3	.15	.11	.15	.16	.08	.12
4	.03	.11	.15	.04	.00	.00
5	.06	.12	.21	-	-	-
Average	.08	.10	.13	.16	.06	.09

Table 21b. The Best Case and Worst Case Analysis for Validity in Predicting Job Proficiency

	Worst Case			Best Case		
Complexity Level	GVN	SPQ	KFM	GVN	SPQ	KFM
1	.52	.52	.25	.60	.52	.35
2	.15	.08	.21	.31	.40	.75
3	.38	.35	.21	.78	.35	.21
4	.31	.26	.12	.69	.54	.52
5	.36	.20	.24	.44	.50	.61
Average	.34	.24	.19	.56	.50	.54

Table 21c. The Best Case and Worst Case Analysis for Validity in Predicting Training Success

	Worst Case			Best Case		
Complexity Level	GVN	SPQ	KFM	GVN	SPQ	KFM
1	.65	.32	.09	.65	.74	.09
2	.29	.26	.01	.71	.26	.25
3	.36	.34	.16	.78	.54	.46
4	.49	.53	.40	.59	.53	.40
5	-	-	-	-	-	-
Average	.37	.33	.14	.74	.47	.38

Table 21 presents the measure of variation in validity from the mean values of Table 18. Table 21a presents the standard deviation of true validity for each level of complexity for job proficiency and for training success. These values are subject to more sampling error than the mean validities, and some of the variation from cell to cell is thus a reflection of this sampling error. The average standard deviation is .10 in comparison to an average mean validity of .40, that is, the standard deviation is about 25 percent of the mean validity.

Table 21b presents the implications of variation in validity for homogeneous hiring for the prediction of job proficiency. For each level of complexity and for each ability, there is a validity distribution. The mean of that distribution is given in Table 18 and the standard deviation is given in Table 21a. From the mean and standard deviation, we can compute the effective range of values that might pertain to a given job from that category. The "worst case" is the 10th-percentile point of the distribution, that is, a value so low that only one in 10 validity values would be lower. The "best case" is the 90th-percentile point of that distribution, that is, a value so high that only one in ten validity values would lie above that value. For example, the average validity of cognitive ability in predicting proficiency of setup work is .56 and the standard deviation is .03. For a specific setup job, the validity might be as low as .52 or it might be as high as .60.

Table 21b shows that even in the worst case, the validity of cognitive ability falls below .31 only for feeding and offbearing jobs. Even for these jobs, the worst case value is .15 which is considerably greater than zero. Thus, Table 21b shows that cognitive ability is a valid predictor of job performance for all jobs.

Table 21b shows similar but somewhat weaker results for perceptual and psychomotor ability. For perceptual ability, the worst case is above .20 for all but feeding and offbearing jobs, and even the value of .08 for

feeding and offbearing jobs is still above zero. For psychomotor ability, the worst case validity is above .21 for all but level 3, and the value of .12 is still well above zero. Thus, Table 21b shows that perceptual and psychomotor ability are valid predictors of job performance in all jobs. If the best predictor is used at each level of complexity, then the worst case validity is less than .31 only for feeding and offbearing jobs, where the worst case value is .21. On the average, the worst case validity for the best single predictor is .34 and the best case validity for the best single predictor is .58.

The focus on the worst case for validity is important for theoretical reasons since it bears on the issue of invalid prediction. The worst case analysis shows that well-constructed general cognitive, perceptual, and psychomotor tests are never invalid. However, the worst case analysis is a very slanted analysis from an applied point of view. The worst case value is deliberately chosen to be an unlikely value. The best case value is just as likely as the worst case value. The most likely values are the mean values shown in Table 18.

Table 21b shows the best case and worst case analysis for the prediction of training success. The worst case for cognitive ability is never less than .29, and the worst case for perceptual ability is never less than .26. However, the worst case for psychomotor ability drops to .01 at level 2. Thus, there are high-complexity jobs where psychomotor has no validity for predicting training success.

At the present time, there is no exact method for obtaining the standard deviation of validity for multiple regression equations such as those recommended in Table 20. However, approximations suggest that the standard deviation would be slightly less than that for the best single predictor. An analysis of variation in validity based on this approximation is shown in Table 22. For job proficiency, the average standard deviation is .09 in comparison to an average validity of .53, that is, the standard deviation is about 17 percent of the mean validity. For training success, the average standard deviation is .15, or about 26 percent of the average validity of .57. For job proficiency, the worst of the worst cases is .21 for feeding and offbearing jobs, while the best of the best cases is .78 for synthesizing and coordinating jobs. For training success, the corresponding range is from .24 to .80.

Table 22

Table 22. An Approximate Analysis of the Variation in Validity Using the Multiple Regression Equations of Table 20 (the Standard Deviation of Validity is Somewhat Overestimated by Using the Standard Deviation of the Best Single Predictor)

Table 22a. Variation in Validity for Multiple Regression Equations Predicting Job Proficiency; i.e., the Regression Equations of Table 20a

Complexity Level	Average Validity	Standard Deviation	Worst Case	Best Case
1	.59	.03	.55	.63
2	.58	.15	.38	.78
3	.53	.15	.33	.73
4	.50	.15	.30	.70
5	.49	.21	.21	.77
Average	.53	.09	.32	.71

Table 22b. Variation in Validity for Multiple Regression Equations Predicting Training Success, i.e., the Regression Equations of Table 20b

Complexity Level	Average Validity	Standard Deviation	Worst Case	Best Case
1	.65	.00	.65	.65
2	.50	.20	.24	.76
3	.59	.16	.38	.80
4	.59	.04	.54	.64
5	-	-	-	-
Average	.57	.15	.37	.77

Variation in Validity: Heterogeneous Hiring

The maximum departure from the mean validity of Table 18 or Table 20 is likely to come under the condition of homogeneous hiring, that is, if an employer is hiring all applicants for exactly the same job. If applicants are being hired for different jobs, then the departure is likely to be much smaller. The reason for this is that the average of several values departs less from the mean than does a single value.

Consider an example. The variation in validity is greatest for feeding and offbearing jobs. If selection is done using the multiple regression equation of Table 20, then the average validity is .49, the standard deviation is .21, and hence the effective range of validity spreads from .21 to .77. If an employer is selecting 100 applicants for exactly the same feeding and offbearing job, then the validity for that particular job could fall anywhere in the feeding-offbearing range. The most likely value is .49, but there is a chance of values as far away as .21 or .77. However, suppose that the employer is hiring people for four different feeding and offbearing jobs, say 25 people for each job. Then the validity for the 100 people hired is the average of the validities for the four jobs. If four numbers are chosen randomly from a distribution, then the average of those four numbers will have the same mean as a single number, but will have a standard deviation that is only half as large as a single number. So if the employer hires 25 people for each of the four feeding and offbearing jobs, then the expected validity is still .49, but the standard deviation is only .10. Thus, the effective range of validity for the heterogeneous employer is from .36 to .62 rather than .21 to .77. If the employer were hiring for 16 feeding and offbearing jobs rather than 4, then the effective range would shrink to .42 to .56. For 49 jobs, the effective range would be from .46 to .52.

The effective range for heterogeneous hiring depends on the exact distribution of persons across jobs. If the number of persons is the same for all jobs, then the standard deviation of validity is reduced by exactly the square root of the number of jobs. If the distribution is uneven, then the formula is more complicated.

Validity Generalization Conclusions

There are now validity generalization studies for over 900 test-job combinations. Two general propositions have arisen from these studies: the validity of reliable cognitive ability tests does not vary much across settings or across times, and most major ability tests have at least some validity for all jobs. Both propositions have been confirmed in the U.S. Employment Service data base.

The results in Table 18 show that validity changes only with very large changes in job content. Consider the extremes in manufacturing jobs. As complexity drops from the highly technical industrial setup jobs to the simplicity of feeding and offbearing jobs, the validity of cognitive ability drops only from .56 to .23. The changes in job content from one organization to another in jobs that have the same DOT code are min ule by comparison.

The results in Table 21 show that cognitive ability is valid in predicting both job proficiency and training success for all jobs. Perceptual ability also predicts both proficiency and training success in all jobs, though with a lower mean and a lower "worst case" than for cognitive ability.

Psychomotor ability is valid for all jobs in predicting job proficiency, but has very low validity in predicting training success in high-complexity jobs and no validity at all for some jobs.

For many years, psychologists have hoped that different jobs would be predicted by different specific cognitive and perceptual aptitudes. However, studies with adequate provision for shrinkage or cross-validation have found little evidence for this. It is a rare job in which multiple regression on specific aptitudes provides any improvement over prediction using general cognitive ability.

The data presented in Table 19 and Table 20 show that differential prediction is very effective if psychomotor ability is considered. In the lower complexity jobs where the validity of cognitive ability falls off, the validity of psychomotor ability is at its highest. Thus, multiple regression equations that vary from one complexity level to another yield a much higher overall level of validity than would be the case for any single predictor. For cognitive ability alone, the average validity is only .45. Using cognitive ability to predict for the three higher complexity levels and psychomotor ability to predict the two lower levels, raises the average validity to .48. Use of combinations in multiple regression equations raises the average validity to .54. Thus, using psychomotor ability in combination with cognitive ability leads to an 18 percent increase in validity.

Even after jobs are stratified by complexity, there is some variation in validity. Some of this variation is due to artifacts, but some may be due to unknown job dimensions--dimensions other than the overall job complexity dimension. However, the variability within categories is not large by practical standards.

CONCLUSION

The U.S. Employment Service data base provides the necessary validation information to design a selection program based on ability for any of the 12,000 jobs in the Dictionary of Occupational Titles. An employer need only know the DOT code for the job in question in order to look up the recommended regression equation in Table 20 and its associated baseline validity. The average validity of the regression equations in Table 20 is .53, a very high value in terms of the utility equations that translate test validity into the improvement in workforce productivity. There is variation in validity within job complexity categories which is assessed in Table 22; but since most employers do heterogeneous hiring, even the worst case provides a highly useful level of validity.

Test use requires not only validity but fairness to minority applicants and economic usefulness as well. There is ample evidence of fairness, under the regression model, for cognitive and perceptual aptitudes (Hunter, Schmidt, &

Rauschenberger, in press). A study using the U.S. Employment Service data base confirms these findings for cognitive and perceptual ability and extends them to psychomotor ability as well. An analysis of fairness and adverse impact for the GATB has been done by Hunter (Note 6).

Maximal improvement in workforce productivity requires two things: High validity and hiring by ranking, either across all applicants or within ethnic groups as allotted by quota. The basis for ranking has been spelled out by Hunter (Note 7) and depends largely on the linearity of the relationship between ability and performance. There is a very large, cumulative, empirical basis for linearity (Hunter & Schmidt, 1982; Schmidt, Hunter, McKenzie, & Muldrow, 1979). A cumulative study of the U.S. Employment Service data base showing linearity was done by Hawk (1970). The study by Hawk shows linearity not only for cognitive and perceptual aptitudes, but for psychomotor aptitudes as well. The empirical work on productivity and performance necessary to use the classic utility formulas was presented in Hunter and Schmidt (1982) and Schmidt, Hunter, McKenzie, and Muldrow (1979). However, a much larger empirical basis is now known (Hunter, & Schmidt, in press; Schmidt & Hunter, in press). A utility analysis specifically geared to the GATB and to the validity generalization analysis presented here has been done by Hunter (Note 5).

Scientific standards for test use, such as the APA Standards (APA, AERA, & NCME, 1974) and the Division 14 Principles (APA, 1980), require that a test be valid if it is to be used for selection. Analysis of the U.S. Employment Service data base shows this requirement to be met for all jobs by the general cognitive, perceptual, and psychomotor abilities. Hunter (Note 4) has shown that the major specific aptitudes have validities parallel to those of the general abilities, although at a lower overall level. Thus, by implication, the Employment Service data base has shown that all reliable, major ability tests are valid for all jobs. That is, the present data show that all major ability tests meet the scientific standards for validity without need for local validation studies.

The Federal Uniform Guidelines on Employee Selection Procedures (USEEOC, USCSC, USDOL, & USDOJ, 1978) make a requirement beyond validity. If a test shows adverse impact, then the test must have higher validity than available alternatives. Hunter and Hunter (Note 8) have recently completed a validity generalization analysis of all common alternatives to ability tests, such as biodata, interviews, training and experience ratings, and so forth. No alternative for entry-level hiring compares favorably with the average validity of .53 shown here for optimal use of cognitive and psychomotor ability tests. Thus, the multiple regression equations of Table 20 also meet the requirements of the Uniform Guidelines without need for local validation studies.

REFERENCE NOTES

1. Droege, R. C. & Boese, R. Development of a generalized aptitude pattern structure with comprehensive occupational coverage. Washington, D.C.: U.S. Employment Service, 1979.

2. Helme, W. E., Gibson, W. A., & Brogden, H. E. An empirical test of shrinkage problems in personnel classification research. (Personnel Research Board Technical Research Note 84). Adjutant General's Office, October 1957.

3. Hunter, J. E. Cumulating results across studies: Correction for sampling error, a proposed moratorium on the significance test, and a critique of current multivariate reporting practices. Invited address presented at the meeting of the American Psychological Association, New York, September 3, 1979. Available from John E. Hunter, Department of Psychology, Michigan State University, East Lansing, MI 48824.

4. Hunter, J. E. The dimensionality of the General Aptitude Test Battery (GATB). Manuscript submitted for publication, 1982.

5. Hunter, J. E. The economic benefits of personnel selection using ability tests: A state of the art review including a detailed analysis of the dollar benefit of the U.S. Employment Service placements and a critique of the low-cutoff method of test use. Washington, D.C.: U.S. Employment Service, 1981.

6. Hunter, J. E. Fairness of the General Aptitude Test Battery (GATB): Ability differences and their impact on minority hiring rates. Washington, D.C.: U.S. Employment Service, 1981.

7. Hunter, J. E. The validity of content valid tests and the basis for ranking. Keynote address presented at the meeting of the International Personnel Management Association Assessment Council, Minneapolis, June 7, 1982.

8. Hunter, J. E. & Hunter, R. F. The validity and utility of alternate predictors of job performance. Washington, D.C.: U.S. Office of Personnel Management, 1982.

REFERENCES

American Psychological Association, Oivision of Industrial-Organizational Psychology. Principles for the validation and use of personnel selection procedures (2nd ed.). Berkeley, CA: Author, 1980.

American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. Standards for educational and psychological tests. Washington, D.C.: American Psychological Association, 1974.

Cain, P. S., & Green, 8. F. The rating of DOT worker functions and worker traits. In A. R. Miller, D. J. Treiman, P. S. Cain, and P. A. Roos (Eds.), Work, Jobs, and Occupations. Washington, D.C., National Academy Press, 1980, 315-335.

Callender, J. C. & Osburn, H. G. Development and test of a new model for validity generalization. Journal of Applied Psychology, 1980, 65, 543-558.

Fine, S. A., & Heinz, C.A. The functional occupational structure. Personnel and Guidance Journal, 1958, 37, 180-192.

Fine, S. A. A structure of worker functions. Personnel and Guidance Journal, 1958, 37, 180-192.

Hawk, J. Linearity of criterion-GATB aptitude relationships. Measurement and Evaluation in Guidance, 1970, 2, 249-251.

Hunter, J. E. Construct validity and validity generalization. In Construct validity in psychological measurement. Princeton, NJ: Educational Testing Service, 1980.

Hunter, J. E., & Schmidt, F. L. The economic benefits of personnel selection using psychological ability tests. Industrial Relations, in press.

Hunter, J. E., & Schmidt, F.L. Fitting people to jobs: The impact of personnel selection on national productivity. In M.D. Dunnette and E. A. Fleishman (Eds.), Human capability assessment. Hillsdale, NJ: Lawrence Erlbaum Associates, 1982.

Hunter, J. E., Schmidt, F. L., & Jackson, G. 8. Meta-analysis: Cumulating results across studies. Beverly Hills, CA: Sage Publications, 1982.

Hunter, J. E., Schmidt, F. L., & Rauschenberger, J. M. Methodological, statistical, and ethical issues in the study of bias in psychological tests. In C. R. Reynolds, (Ed.), Perspectives on bias in mental testing. New York: Plenum Press, in press.

King, L. M., Hunter, J. E., & Schmidt, F. L. Halo in multidimensional forced choice performance evaluation scale. Journal of Applied Psychology, 1980, 65, 506-516.

Lent, R. H., Aurbach, H. A., & Levin, L. S. Research design and validity assessment. Personnel Psychology, 1971, 24, 247-274.

McCormick, E. J., Jeanneret, P. R., & Mecham, R. C. A study of job characteristics and job dimensions as based on the Position Analysis Questionnaire (PAQ). Journal of Applied Psychology Monograph, 1972, 56, 347-368.

Mecham, R. C., McCormick, E. J., & Jeanneret, P. R. Position Analysis Questionnaire: Users Manual (System II). Logan, Utah: PAQ Services, Inc., 1977.

Pearlman, K. The Bayesian approach to validity generalization: A systematic examination of the robustness of procedures and conclusions. Unpublished doctoral dissertation, George Washington University, 1982.

Pearlman, K., Schmidt, F. L., & Hunter, J. E. Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. Journal of Applied Psychology, 1980, 65, 374-406.

Pearlman, K., Schmidt, F. L., & Hunter, J. E. Effects of alternate job grouping methods on selection procedure validity and validity generalization. Journal of Applied Psychology, in press.

Schmidt, F. L., Gast-Rosenberg, I., & Hunter, J. E. Validity generalization results for computer programmers. Journal of Applied Psychology, 1980, 65 643-661.

Schmidt, F. L., & Hunter, J. E. Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 1977, 62, 529-540.

Schmidt, F. L., & Hunter, J. E. Moderator research and the law of small numbers. Personnel Psychology, 1978, 32, 215-232.

Schmidt, F. L., & Hunter, J. E. An empirical test of estimates of employee variability in productivity derived from studies of selection procedure utility. Journal of Applied Psychology, in press.

Schmidt, F. L., Hunter, J. E., & Caplan, J. R. Validity generalization results for two job groups in the petroleum industry. Journal of Applied Psychology, 1981, 66, 261-273.

Schmidt, F. L., Hunter, J. E., McKenzie, R. C. and Muldrow, T. The impact of valid selection Procedures on workforce productivity. Journal of Applied Psychology, 1979, 64, 609-626.

Schmidt, F. L., Hunter, J. E., & Pearlman, K. Task differences as moderators of aptitude test validity in selection: a red herring. Journal of Applied Psychology, 1981, 66, 166-185.

Schmidt, F. L., Hunter, J. E., & Pearlman, K. Progress in validity generalization: Comment on Callender and Osburn and further developments. Journal of Applied Psychology, in press.

Schmidt, F. L., Hunter, J. E., Pearlman, K., & Shane, G. S. Further tests of the Schmidt-Hunter Bayesian validity generalization procedure. Personnel Psychology, 1979, 32, 257-281.

U.S. Department of Labor. Dictionary of Occupational Titles (4th ed.). Washington, D.C.: U.S. Government Printing Office, 1977.

U.S. Department of Labor. Guide for Occupational Exploration. Washington, D.C.: U.S. Government Printing Office, 1979.

U.S. Equal Employment Opportunity Commission, U.S. Civil Service Commission, U.S. Department of Labor, & U.S. Department of Justice. Uniform guidelines on employee selection procedures. Federal Register, 1978, 43(166), 38295-38309.

U.S. Government Printing Office 1984 - 421-608/4641